

Capstone Project – The Battle of Neighborhoods

1. Introduction

The Gym, Health and Fitness Clubs industry in New York has benefited in recent years from increased awareness about the benefits of exercise. Over the five years to 2018, powered by strong growth in per capita disposable income, industry revenue is expected to increase at an annualized rate of 0.8% to reach \$3.1 billion, including an increase of 1.4% in 2018 alone [1]. In terms of such tremendous market potential, gym industry is attracting more and more value investors to make investments, especially in the New York City("NYC"), which is the most populous city with the largest population in the United States.

Before entrepreneurs to start their business in the Gym, Health and Fitness Clubs industry, they need to find out where they would open it? What would they consider when selecting a location? By exploring gyms, fitness clubs in major parts of New York City, this project is to figure out whether the neighborhood of gyms and fitness clubs is an essential factor for the success of a gym with sound analysis.

2. Data

For this project, the following data was utilized:

- **New York City geographical data:** contains information of Boroughs, Neighborhoods along with their latitude and longitude. This dataset is used to explore the neighborhoods of New York City and exists for free on the web (https://cocl.us/new_york_dataset).
- **Foursquare API:** by using this API, we can get all the venues in each neighborhood and filter these venues to get gym and fitness clubs.

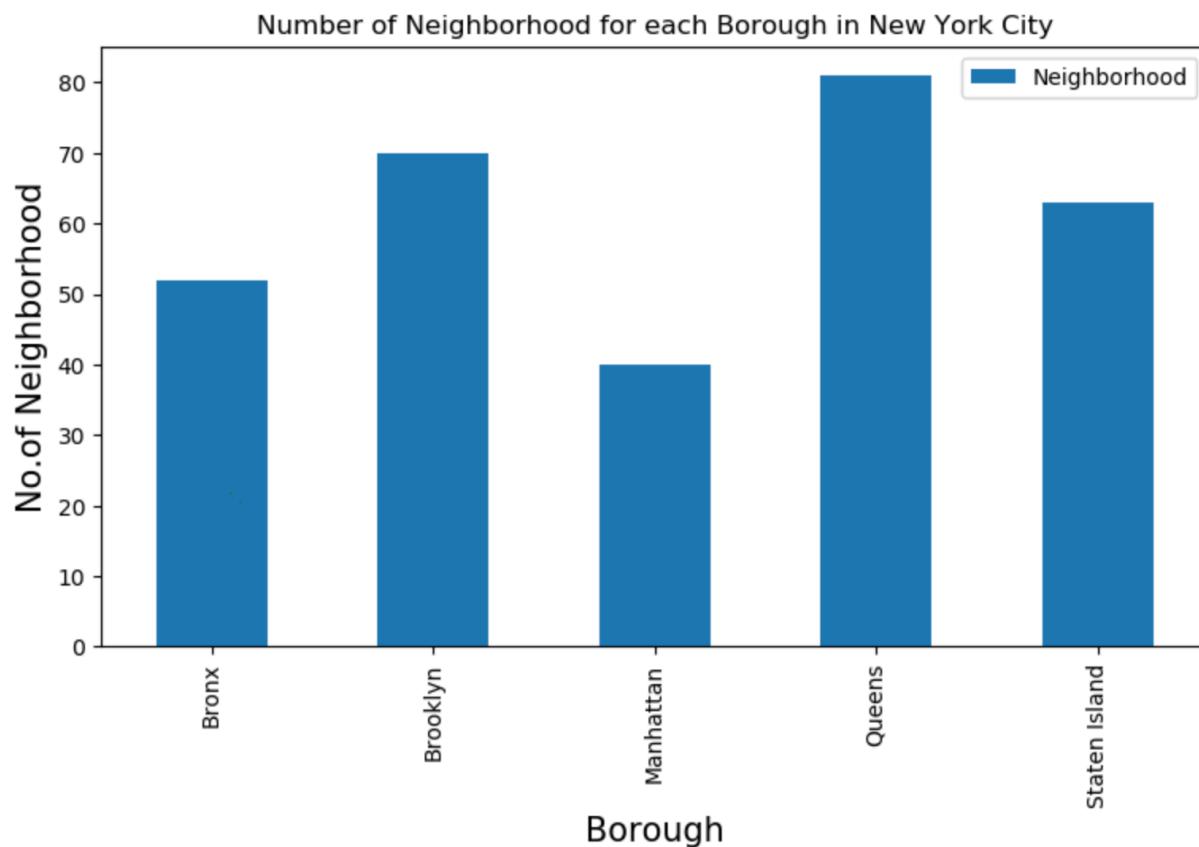
3. Methodology with Results

3.1 Descriptive data analysis for New York City Neighborhood data

First of all, let's get the response geographical data of NYC online, converting it to a Python object, then load the data to a pandas dataframe including Borough name, Neighborhood name, Neighborhood Latitude and Neighborhood Longitude for the descriptive data analysis. The first 10 rows of the dataframe is as following:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

Next, let's visualize the number of neighborhoods for each Borough in New York City.



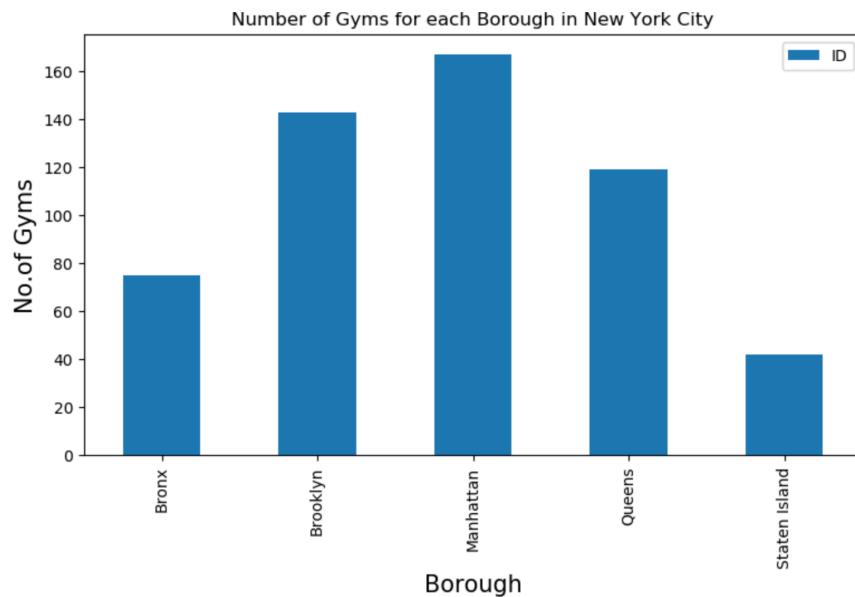
It shows that Queens includes largest number of Neighborhoods. It may raise the question that *does Queens have the largest number of gyms?* Let's solve this in the later section. So far, we have explored the New York City Neighborhood data, getting the geographical coordinates which is going to be used to find the nearby gyms by interacting with Foursquare API.

3.2 Descriptive data analysis for Gym venues of each neighborhood

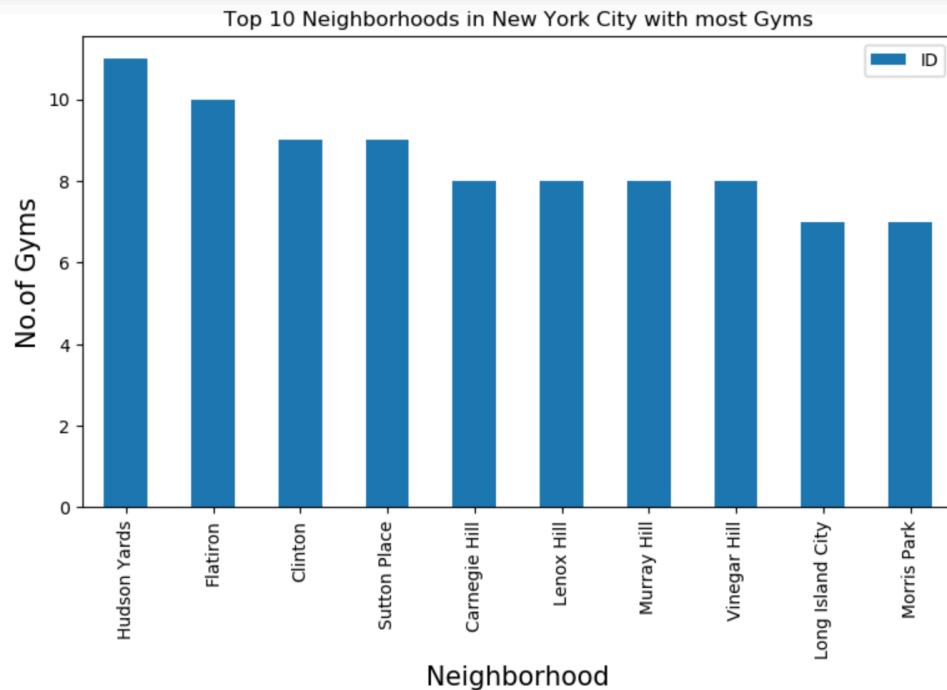
In this section, we will explore the gym venues of each neighborhood by interacting with Foursquare API, collecting the name, ID, latitude, longitude of each gym, given the geographical coordinates we got for each neighborhood in above section.

Borough	Neighborhood	ID	Name	Latitude	Longitude
Bronx	Bedford Park	532ad23e498e87bb1035acf5	Crunch	40.872635	-73.874847
Bronx	Bedford Park	5cf7f63cb9b37b002c7b4d0f	Blink Fitness	40.873893	-73.888768
Bronx	Bedford Park	4c5ee8ae7f661b8d94464f1c	APEX - Lehman College	40.874849	-73.893296
Bronx	Belmont	52af0e3c11d2fb9a346ce0f7	Blink Fitness	40.859373	-73.898316
Bronx	Belmont	52af0dc711d21e1625464d63	Blink Fitness	40.859482	-73.898520
Bronx	Bronxdale	4c518ceb9426c9283e19f373	Powerhouse Gym	40.844493	-73.857738
Bronx	Castle Hill	4c38a7920a71c9b6d30241c9	Bronx YMCA	40.812425	-73.846126
Bronx	Claremont Village	5632f45c498ef690c483ecd4	Retro Fitness	40.831190	-73.905395
Bronx	Clason Point	4c38a7920a71c9b6d30241c9	Bronx YMCA	40.812425	-73.846126
Bronx	Co-op City	4c2a235d97d00f47fa7e42ea	24 Hour Fitness	40.867818	-73.824984

We can visualize the number gyms of each Borough and each neighborhood as following:



It shows that Manhattan has the largest number of gyms. This answers the question we have at the end of section 3.1 that, although Queens has largest number of neighborhoods, it doesn't have largest number of gyms. Similarly, we can visualize the top 10 neighborhoods that have most gyms.



After exploring the distribution of gyms by Borough and Neighborhood, we could ask question like, *what could impact location of a gym?* or *what kind of neighborhood does a gym owner would like to choose to open the business?* This is also the purpose of this entire project to help entrepreneurs to find a right location to business in the Gym, Health and Fitness Clubs industry.

To answer this question, we will focus on the good rating gyms (rating ≥ 9.0), to see what the common nearby venues are surrounding the good rating gyms. Next we will add more gym venue details like rating, likes and tips from Foursquare API.

Manhattan is the most densely populated borough in New York City, and that population density carries over to offices and co-working spaces. There is an incredibly high demand for gym/fitness spots in Manhattan. For entrepreneurs to start gym business, **we will look into the gyms in Manhattan in the following analysis.**

3.3 Find the good gyms in Manhattan

In order to find the “good gyms” in Manhattan, we need to pull the gym details including rating, number of likes, number of tips from Foursquare, given the venue ID of each gym.

Borough	Neighborhood	ID	Name	Latitude	Longitude	Rating	Likes	Tips	
0	Manhattan	Battery Park City	541a2341498e51ce1037963a	Equinox Brookfield Place	40.712704	-74.014995	9.0	174	14
1	Manhattan	Battery Park City	4a6e331af964a52031d41fe3	Equinox Tribeca	40.714099	-74.009686	8.8	250	21
2	Manhattan	Battery Park City	58a253c830ecc66c9e5b40a0	The Class by Taryn Toomey	40.712753	-74.008734	9.1	19	2
3	Manhattan	Battery Park City	5a697271ad178907f79be157	Crunch - FiDi	40.708614	-74.010013	8.7	20	1
4	Manhattan	Battery Park City	4a2f3322f964a520b6981fe3	Equinox Wall Street	40.707273	-74.010543	8.8	311	50
5	Manhattan	Carnegie Hill	574dc71c498ef19a1c178913	Equinox East 92nd Street	40.782500	-73.950580	9.3	72	12
6	Manhattan	Carnegie Hill	5158e6bde4b079b32ba7e7fd	Crank NYC	40.783209	-73.950862	8.8	48	9
7	Manhattan	Carnegie Hill	4c9a873ad4b1b1f79dff035	92nd Street Y May Center	40.783108	-73.952758	8.5	74	10
8	Manhattan	Carnegie Hill	52b9c263498ebbf8b240197	ChaiseFitness	40.783065	-73.952648	8.3	22	3
9	Manhattan	Carnegie Hill	4a97e23df964a520db2920e3	Equinox East 85th Street	40.778001	-73.954143	8.8	207	30
10	Manhattan	Carnegie Hill	4b9d849ff964a52087b136e3	Club Normandie	40.784921	-73.948533	8.5	32	13

There could be many standards to define if a gym business is successful/good. One of the most common factors is online rating by customers. We also collected the number of likes and number of Tips which could reflect how popular it is among the customers. Let's look at the correlation among the three factors: rating, number of likes and number of tips:

	Rating	Likes	Tips
Rating	1.000000	0.288911	0.305661
Likes	0.288911	1.000000	0.608718
Tips	0.305661	0.608718	1.000000

It doesn't show the significant relation between Rating and Likes, Rating and Tips. While in this project, we define a good/successful gym business with Rating ≥ 8.5 , with at least 50 likes and at least 1 Tip.

Gym_ID	Gym_name	Gym_lat	Gym_lng	Rating	Likes	Tips
541a2341498e51ce1037963a	Equinox Brookfield Place	40.712704	-74.014995	9.0	174.0	14.0
4a6e331af964a52031d41fe3	Equinox Tribeca	40.714099	-74.009686	8.8	250.0	21.0
4a2f3322f964a520b6981fe3	Equinox Wall Street	40.707273	-74.010543	8.8	311.0	50.0
574dc71c498ef19a1c178913	Equinox East 92nd Street	40.782500	-73.950580	9.3	72.0	12.0
4c9a873ad4b1b1f79dff035	92nd Street Y May Center	40.783108	-73.952758	8.5	74.0	10.0
4a97e23df964a520db2920e3	Equinox East 85th Street	40.778001	-73.954143	8.8	207.0	30.0
4d2cc14bb47d2c0fd1608594	Barry's Bootcamp	40.741647	-73.995769	9.2	443.0	64.0
4a72e5fef964a52050db1fe3	Chelsea Piers Fitness	40.746645	-74.010057	8.8	493.0	100.0
46b78e2cf964a520bc491fe3	Manhattan Plaza Health Club (MPHC)	40.759967	-73.994585	8.7	108.0	34.0
4eb5c125f9f40c39c596bb4a	Mark Fisher Fitness - Ninja Clubhouse	40.756836	-73.994719	9.2	72.0	14.0

3.4 Explore the nearby venues for each good gym

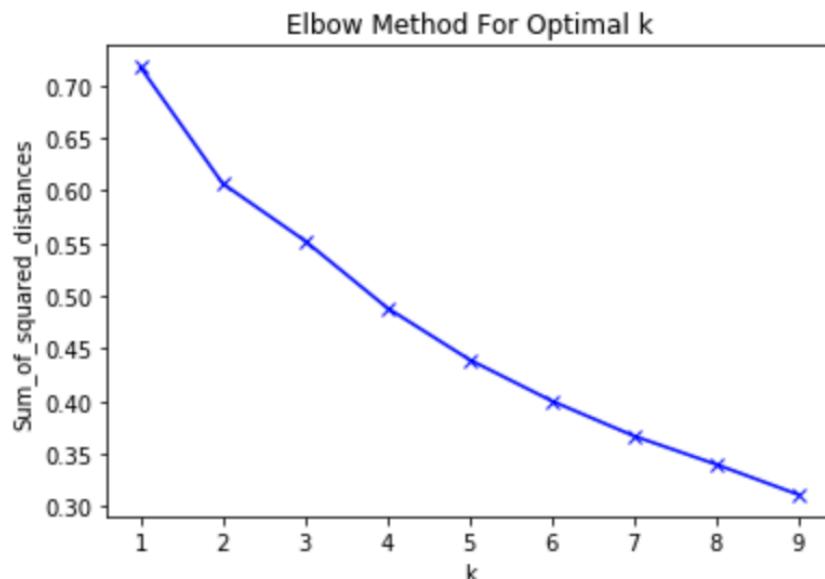
We explored the venues within 500 meters for each good gym. It's important to find the most common venues surrounding the good gyms, which could indicate the neighborhood characters for a successful gym business. We find the 5 most common venues around each gym:

	Gym ID	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	42af6f80f964a5205a251fe3	Equinox East 44th Street	Coffee Shop	Hotel	Japanese Restaurant	Grocery Store	Sandwich Place
1	45938647f964a52045401fe3	Equinox Sports Club New York	Italian Restaurant	Theater	Plaza	Concert Hall	Café
2	46b78e2cf964a520bc491fe3	Manhattan Plaza Health Club (MPHC)	Coffee Shop	Gym / Fitness Center	Wine Shop	Gym	Theater
3	49bd4e25f964a52067541fe3	Equinox SoHo	Italian Restaurant	Mediterranean Restaurant	Café	Clothing Store	Cosmetics Shop
4	49c9439bf964a5204e581fe3	Equinox Greenwich Avenue	Italian Restaurant	Coffee Shop	Bakery	Park	Café
5	49fa353af964a520da6d1fe3	Crunch Fitness - 34th Street	Coffee Shop	Bar	Café	Mexican Restaurant	Pizza Place
6	49fce426f964a520ed6e1fe3	Equinox Columbus Circle	Concert Hall	Jazz Club	Hotel	Bakery	Performing Arts Venue
7	4a2f3322f964a520b6981fe3	Equinox Wall Street	Coffee Shop	Pizza Place	Hotel	Cocktail Bar	Falafel Restaurant
8	4a68fb35f964a52064cb1fe3	Equinox East 63rd Street	Italian Restaurant	Hotel	Gym / Fitness Center	Boutique	Coffee Shop
9	4a6e331af964a52031d41fe3	Equinox Tribeca	Coffee Shop	Hotel	Gym	Bakery	Memorial Site

At first glance, it seems that Café, Bakery, and some other light food store is very common around these gyms. It makes sense that people would like to get some light food before or after the workout. We need to further investigate the cluster of these gyms to find pattern of clusters.

3.5 Cluster Gyms: K-means Clustering Analysis

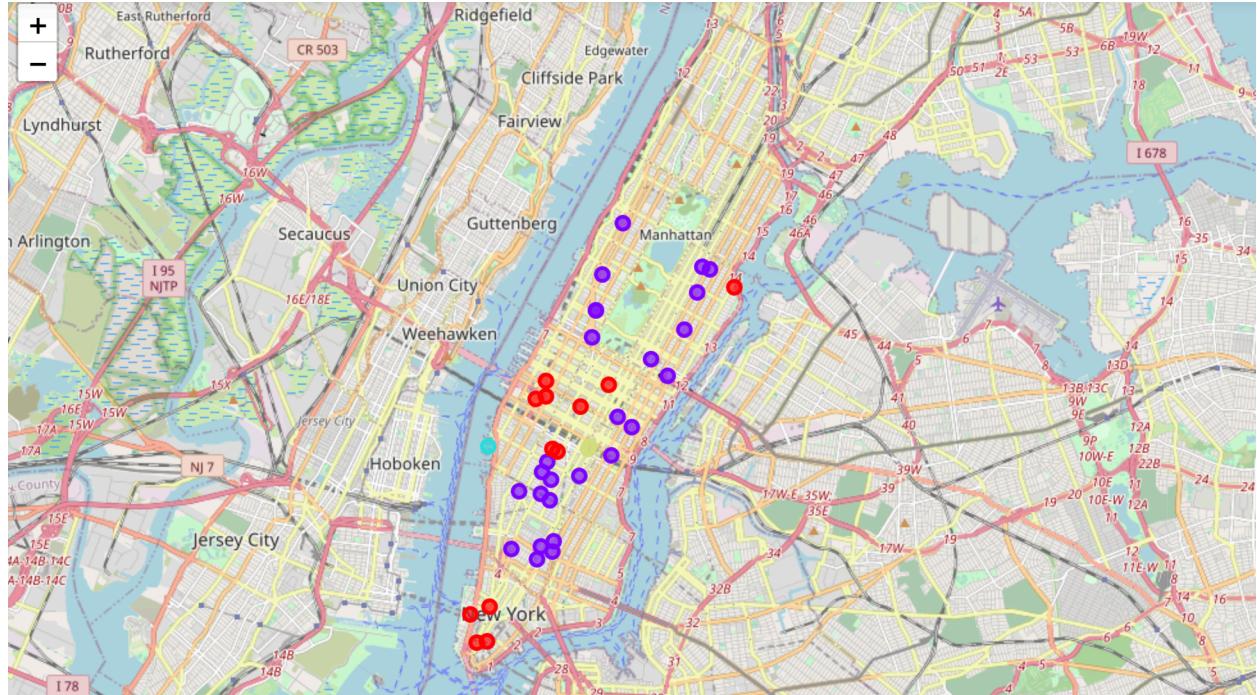
We draw the *elbow curve* to find the optimal K value for the k-mean clustering analysis.



There is no clear 'elbow' shown on the chart, here we chose 4 as the K to perform the clustering analysis. The k-means clustering can help us do the clustering for those good gyms, which could help us to see the common surrounding more clearly of each cluster.

Gym ID	Gym_Lat	Gym_Lng	Cluster Labels	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
541a2341498e51ce1037963a	40.712704	-74.014995	0	Equinox Brookfield Place	Park	Hotel	Coffee Shop	Wine Shop	Gym
4a6e331af964a52031d41fe3	40.714099	-74.009686	0	Equinox Tribeca	Coffee Shop	Hotel	Gym	Bakery	Memorial Site
4a2f3322f964a520b6981fe3	40.707273	-74.010543	0	Equinox Wall Street	Coffee Shop	Pizza Place	Hotel	Cocktail Bar	Falafel Restaurant
574dc71c498ef19a1c178913	40.782500	-73.950580	1	Equinox East 92nd Street	Coffee Shop	Pizza Place	Gym	Café	Japanese Restaurant
4c9a873ad4b1b1f79dff035	40.783108	-73.952758	1	92nd Street Y May Center	Coffee Shop	Pizza Place	Gym	Café	Wine Shop
4a97e23df964a520db2920e3	40.778001	-73.954143	1	Equinox East 85th Street	Italian Restaurant	Pizza Place	Bar	Bakery	Coffee Shop

We can see the clusters on the map:



Next, let's examine the Clusters and determine the discriminating venue categories that distinguish each cluster.

Cluster 1

	Gym ID	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	541a2341498e51ce1037963a	Equinox Brookfield Place	Park	Hotel	Coffee Shop	Wine Shop	Gym
1	4a6e331af964a52031d41fe3	Equinox Tribeca	Coffee Shop	Hotel	Gym	Bakery	Memorial Site
2	4a2f3322f964a520b6981fe3	Equinox Wall Street	Coffee Shop	Pizza Place	Hotel	Cocktail Bar	Falafel Restaurant
8	46b78e2cf964a520bc491fe3	Manhattan Plaza Health Club (MPHC)	Coffee Shop	Gym / Fitness Center	Wine Shop	Gym	Theater
9	4eb5c125f9f40c39c596bb4a	Mark Fisher Fitness - Ninja Clubhouse	Theater	Coffee Shop	Gym / Fitness Center	Hotel	Dance Studio
10	533629bf498ed70d66886dd6	TITLE Boxing Club NYC - Midtown West	Hotel	Theater	Coffee Shop	Gym / Fitness Center	Cocktail Bar
11	568dd8f0498e49b9d660021b	Physique 57	Park	Coffee Shop	Hotel	Pizza Place	Mexican Restaurant
17	4efe614e0e618ea7655927b8	Body & Pole	Coffee Shop	Hotel	Gym / Fitness Center	Hotel Bar	Spa
28	53851395498e2379f6178a0d	Equinox Bryant Park	Theater	Hotel	Coffee Shop	Bakery	Mediterranean Restaurant
30	4a705a19f964a5204ed71fe3	Equinox	Hotel	Indian Restaurant	Art Museum	Food & Drink Shop	Plaza
31	4f038a38e30090dfdb03701c	iLoveKickboxing	Gym / Fitness Center	Coffee Shop	Hotel	Hotel Bar	Yoga Studio
40	4a6f77a1f964a5203ed61fe3	Asphalt Green	Coffee Shop	Gym	Park	Pizza Place	Mexican Restaurant

Cluster 2

	Gym ID	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	574dc71c498ef19a1c178913	Equinox East 92nd Street	Coffee Shop	Pizza Place	Gym	Café	Japanese Restaurant
4	4c9a873ad4b1b1f79dff035	92nd Street Y May Center	Coffee Shop	Pizza Place	Gym	Café	Wine Shop
5	4a97e23df964a520db2920e3	Equinox East 85th Street	Italian Restaurant	Pizza Place	Bar	Bakery	Coffee Shop
6	4d2cc14bb47d2c0fd1608594	Barry's Bootcamp	Coffee Shop	Gym / Fitness Center	Grocery Store	Yoga Studio	Cycle Studio
12	54cd0dab498e6e5d8a9d7866	The Fhitting Room	Gym / Fitness Center	Coffee Shop	Gym	Yoga Studio	Italian Restaurant
13	574da377498ef19a1bd08764	Equinox Gramercy	Indian Restaurant	Bar	Café	Thrift / Vintage Store	Italian Restaurant
14	52a9166611d204ca2a31b144	CityRow	Yoga Studio	Gym / Fitness Center	Cosmetics Shop	Vegetarian / Vegan Restaurant	Bookstore
15	4f9e99f7e4b06064311cf82e	Body Space Fitness Inc	Italian Restaurant	Gym / Fitness Center	Arts & Crafts Store	Steakhouse	Tapas Restaurant
16	586ff8a545c3ed0998f3fcfb	Rumble	Gym / Fitness Center	Yoga Studio	Coffee Shop	Grocery Store	Café
18	4a7ca51ff964a52012ed1fe3	Physique 57	Italian Restaurant	Café	Coffee Shop	Indian Restaurant	Burger Joint
19	4b59affef964a520d69228e3	Work Train Fight	Italian Restaurant	Clothing Store	Yoga Studio	Gym / Fitness Center	Spa
20	49bd4e25f964a52067541fe3	Equinox SoHo	Italian Restaurant	Mediterranean Restaurant	Café	Clothing Store	Cosmetics Shop
21	54bee7a6498ebaaf4ebe9d13	Overthrow Boxing NYC	Italian Restaurant	Gift Shop	Coffee Shop	Yoga Studio	Cocktail Bar

Cluster 3

	Gym ID	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	4a72e5fef964a52050db1fe3	Chelsea Piers Fitness	Art Gallery	Park	Italian Restaurant	Bus Station	Coffee Shop

Cluster 4

```
d.loc[good_gym_merged['Cluster Labels'] == 3, good_gym_merged.columns[[0,4] + list(range(5, good_gym_merged.shape[1]))]]
```

Gym ID	Gym Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
32 5672e6a5498ea77d7c6aafe4	305 Fitness MIDTOWN	Korean Restaurant	Bakery	Gym / Fitness Center	Spa	Café
33 544993c4498e8104085787de	CrossFit Solace	Korean Restaurant	Gym / Fitness Center	Spa	Bakery	Japanese Restaurant

4. Discussion

As a recommendation to those who plan to operate a gym in Manhattan, location selection is only one fundamental problem to think over. It cannot solve the problem of whether a gym could make good revenue but may give us indication that what kind of venues are surrounding a popular gym with good ratings. And as for location suggestion, it offers an opportunity analysis but lacks risk analysis, like the cost of the location and competition in that area.

Although in this report, it demonstrates the relations between location and ratings, but ratings might not reflect the operation status of the gym. A gym with a high rating could still be unprofitable, which is unsuccessful from a business perspective. So, the suggestion is relatively narrow. To suggest more practical and profitable ideas, the relationship between customer reactions and financial performance should be evaluated.

With all these analyses done, the report finally becomes constructive for a gym owner in the real business world.

5. Conclusion

In this Capstone project, we're trying to give the gym owner or entrepreneurs who want to step in gym business in Manhattan a hint that where could be good location of gym, surrounded by the light food store like bakery, coffeeshop, pizza place is a location worth further investigation.